DOCUMENT RESUME

ED 411 310 TM 027 365

AUTHOR Solano-Flores, Guillermo; Raymond, Bruce; Schneider, Steven A.

TITLE Computer-Assisted Portfolio Scoring: Can Technology Enhance the Process of Scoring Portfolios?

SPONS AGENCY National Board for Professional Teaching Standards, Detroit, MI.

PUB DATE 1997-03-00

NOTE 20p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).

PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS Art Teachers; *Computer Assisted Testing; Computer Software; Computer Software Evaluation; Educational Technology; Elementary Secondary Education; *Evaluators; Field Tests; Holistic Evaluation; *Interrater Reliability; *Portfolio Assessment; Portfolios (Background Materials); *Scoring

IDENTIFIERS *National Board for Professional Teaching Standards

ABSTRACT

The need for effective ways of monitoring the quality of scoring of portfolios resulted in the development of a software package that provides scoring leaders with updated information on their assessors' scoring quality. Assessors with computers enter data as they score, and this information is analyzed and reported to scoring leaders. The developed structural scoring approach, PerformAce, was tested with responses of teachers to the portfolio component of a teacher certification assessment developed by the National Board for Professional Teaching Standards (NBPTS). Assessors used the PerformAce software as an electronic scoring form with screens that display the description of an attribute at different levels of performance attached to a scale. When all attributes are scored, the program asks assessors for a holistic score for the examinee's performance. Both attribute and holistic scores are stored in the assessor's computer and then transferred to a central computer to allow scoring leaders to identify patterns of ineffective or inefficient scoring. The approach was used with the field test of an NBPTS art teacher examination with eight portfolio entries and four assessment center exercises. Assessors were 127 art teachers who rated 928 portfolio responses. The best possible estimate of the reliability of the assessment was 0.81, and using computers did not affect this reliability. Almost all of the assessors (94%) agreed that the approach helped them make informed decisions and 73% of the scoring leaders thought the software helped them preserve the quality of scoring. Results suggest that the computer-assisted scoring approach does not have an adverse impact on interrater reliability and that computers can be useful tools to manage scoring sessions. (Contains 1 table, 4 figures, and 12 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL
HAS BEEN GRANTED BY
Guillermo
Solano-Flores
TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

ED 411 310

Computer-assisted Portfolio Scoring: Can Technology Enhance the Process of Scoring Portfolios?

Guillermo Solano-Flores
Bruce Raymond
Steven A. Schneider

TM027365

This paper is prepared for the Annual Meeting of the American Educational Research Association in Chicago, IL

BEST COPY AVAILABLE

Computer-assisted portfolio scoring: Can technology enhance the process of scoring portfolios?[1]

Guillermo Solano-Flores, Bruce Raymond, & Steven A. Schneider
WestEd

Paper presented at the Annual Meeting of the American Educational Research Association Chicago, IL, April 24-28, 1997

As educators and institutions turn their attention to portfolios (e.g., Baratz-Snowden, 1991; Haertel, 1990; Schneider & Austin, 1994; Shulman, 1989; Wolf, 1991) as alternative methods for identifying accomplished teachers with certification purposes, serious cost-efficiency (e.g., Baratz-Snowden, 1991; Reckase, 1995) and interrater reliability (e.g., Koretz, Stecher, Klein, McCaffrey, 1994) issues arise.

One of the challenges posed by portfolios is inherent to the traditional concern about the viability of judging complex performance in an assessment--interrater reliability (cf. Fitzpatrick & Morrison, 1971). Examinee's responses must be independently scorerd by at least to assessors and a third assessor if they have a considerable disagreement (see Wolf, 1994). The scoring of a candidatate's response to a single portfolio entry (exercise) may take up to 45 minutes; it may involve reviewing a collection of products such as a narrative several pages long, a videotape of the examinee's teaching, and some supporting documentation or

---

[1] Funding for the development of the software used in the investigation was provided by WestEd. This investigation was carried out with support and funding from the National Board for Professional Teaching Standards. We wish to thank Dean Nafziger, Don Barfield, Phil Kearney, Jim Smith, and Blair Gibb for the different kinds of support they provided; Lloyd Bond, Dick Jaeger, Lee Cronback, Ed Haertel, and Rich Shavelson for their comments on the scoring approach and the computer software; Sue Austin, Kirsten Daehler, and Jerome Shaw for their comments on the usability of the software; and Marnie Thompson for her participation in the project. Especially, we wish to thank Joan Peterson, Mike Timms, Jody McCarthy, Kirsten Daehler, Kim O'Neill, and Harriet Kossman, for their full support, commitment, and creativity, which made possible the development and use of the scoring approach and the software; and the art teachers who enthusiastically participated as assessment developers or assessors.

samples of student work. Consequently, scoring portfolios is more expensive and time consuming than scoring other types of assessment.

Becuase not all individals can be trained to score performance consistently, assessors must be continually monitored and re-calibrated (cf. Wigdor & Green, 1991). In a widely used scoring procedure, a scoring leader reviews the performance of each assessor by examining how frequently and how much this assessor's scoring disagrees with other assessors' scoring. This may involve "reading behind" the responses scored by that assessor to check if the scores assigned are reasonable. Some assessors need to be re-calibrated by having them review "benchmark" responses or get additional training. In some cases, assessors who do not perform well should not participate in scoring and must be dismissed because they introduce too much measurement error and because too many "third scorings" is costly and jeopardizes the completion of the scoring.

Although re-calibration and dismissal decisions are to a great extent the core of a good management of the scoring sessions, they are often made quickly on a subjective basis. For example, the scoring leader may not have the time to do all the "read-behinds" needed to support these decisions.

The need for effective approaches to monitoring the quality of scoring made us develop a software package that provides scoring leaders with updated information on their assessors' scoring quality. Assessors are provided with computers, so they enter data as they score; this information is analyzed and reported to scoring leaders. In addition to helping scoring leaders make informed decisions on their assessors' scoring quality, this approach can potentially contribute to reduced scoring time: available evidence (Cedeño & Ruiz-Primo, 1982) indicates that the average time used to score complex responses can be drastically reduced if assessors are provided with appropriate software for computer-assisted scoring.

In this paper we describe the characteristics of the software and report the results we obtained when we used this approach to score portfolios in a teacher certification assessment. We discuss how the formal properties of the scoring approach allowed us to design that software. Then we describe how the software was used to score the responses of teachers in the portfolio component of a teacher certification assessment. Next, we evaluate the effectiveness--ability to accomplish the intended scoring goals--and efficiency--ability to accomplish those goals with ease and at a low cost--of this approach. Finally, we discuss possible ways to enhance computer-assisted scoring.

Assessment Context

In 1986, a new organization was formed to create a national system for the voluntary certification of teachers, the National Board of Professional Teaching Standards (NBPTS). Its goal is to establish advanced professional standards for teaching and move away from sole reliance on traditional paper-and-pencil tests of teaching toward more complex forms of assessment. As a result of the Board's work, more sophisticated forms of assessment have been developed with the idea of capturing a wider range of teacher knowledge and skill than do paper-and-pencil tests. These new forms of assessments, that include portfolios and assessment center exercises, attempt to capture the interactive and adaptive nature of teaching work that requires decision making, judgment, reflection, and the sensitive management of multiple dilemmas of the teaching practice.

Scoring Approach

To assure that decisions about performance be based on an appropriate level of analysis, neither too global nor fragmentary (see Haertel, 1990), we developed a design that provided assessors with substantial guidance on all aspects that required performance evaluation. We designed the structural scoring approach to make detailed judgments on complex performance without giving up the capability of

evaluating performance as a whole. In this approach, performance is specified as a set of attributes--aspects of performance. As an example of the detail provided to assessors, Figure 1 shows the information for a single attribute, Scope of Goal, one of the many considered in scoring a portfolio entry. The description specifies what performance looks like at four levels of quality.

The arrangement of all the attributes for a portfolio entry or an assessment center exercise gives as a result a matrix like that illustrated in Figure 2 (the font size of the text has been intentionally reduced). This matrix describes performance at two levels of specificity. At a specific level, a row provides a detailed description of a certain attribute at different levels of quality; at a global level, a column provides a description of performance as a whole for a particular level of quality. The structural approach, then, allows to score performance both at the attribute level and as a whole.

Computer-Assisted Scoring Software Package

We took advantage of the formal properties of the structural scoring approach and developed PerformAce™, a computer program for the use of assessors. This program stores the text that goes in each cell in the matrix (see Figure 2) and displays on the computer monitor the information that assessors need to see to make their scoring decisions. More specifically, the text for all attributes is stored in a database keyed to the entry and attribute, which allows to easily change descriptive text during the development phase. In fact, the entire set of attributes can be interchanged with a new set of descriptors without rewriting the program--the display of attributes is independent of their content and structure.

Assessors use PerformAce™ as an electronic scoring form. They work with "screens," instead of, or in addition to handling rubrics printed on paper (Figure 3). Each screen displays the description of an attribute at different levels of performance attached to a scale. To score the examinee's performance on that attribute, the

assessor selects the description that best fits the characteristics observed in performance; then assigns a score to the attribute by clicking the mouse on the box for the appropriate scale point. Once the attribute is scored, the screen for the next attribute appears on the monitor and the same process is repeated. During this phase of scoring at the attribute level, the program allows assessors to review and change the scores. When all attributes have been scored, the assessor is asked to give a holistic score to the examinee's performance as whole. Once holistic scoring begins, the attribute scores are no longer accessible and cannot be reveiwed or changed during the holistic scoring phase.

As scoring takes place, the attribute and holistic scores are entered and stored in the assessor's computer. When the assessor has finished scoring the response, the information can be transferred into a central computer for analysis. For each response scored by an assessor A on a specific portfolio entry, the central computer keeps a cumulative record of the following variables, printed in what, from now on, we call, monitoring form: (1) ID number of the candidate whose response was scored, (2) sequence of scoring--whether A was the first, second, or third assessor to score the response, (3) scoring time--time used to score the response, (4) structural score--average of the attribute scores, (5) standard deviation of the attribute scores, (6) interrater reliability--correlation coefficient between the attribute scores assigned by assessors A and B, (7) holistic score, (8) adjudication--whether the difference between the holistic scores given by assessors A and B exceeds certain pre-specified value and a third assessor is needed to score that response, (9) adjudication won?--whether the holistic score assigned by A was closer to the score assigned by C than the score assigned by B was, and (10) holistic scores given by assessors A, B, and C (the holistic scores given by assessor A are underlined).

Some of these variables can be used to evaluate the quality of scoring. For example, a consistently small standard deviation in the attribute scores given by a

particular assessor to a number of examinees suggests a tendency to assign only scores within a limited range. Although none of these indicators is sufficient to make an accurate judgment on the assessors' performance, used appropriately and in combination they allow to identify patterns of ineffective or inefficient scoring and to make informed decisions on the assessors (e.g., dismissal, re-calibration) and to provide assessors with specific feedback.

Knowing that the computer-literacy of assessors ranged widely, we designed Performace™ under the assumption that assessors were completely unfamiliar with computers and insured that the program would be used with only minimal training by such unskilled users--testing and reviewing the program countless times for characteristics such as ease of use, user-resiliency and robustness to accidents (e.g., hitting the wrong button) or intentional damage caused by "messing around" with the keyboard.

Method

We used the scoring approach and the software for computer-assisted scoring during the field test scoring sessions for the NBPTS Visual Arts, Early Adolescence Through Young Adulthood Assessment held at San Francisco by WestEd in July of 1996. The assessment consisted of 8 portfolio entries and 4 assessment center exercises. The portfolio entries had been completed by the candidates throughout a period of 6 months and addressed art teaching skills; the types of products submitted by the examinees included videotapes of their own teaching, narratives of their work, and samples of student work. The assessment center exercises were paper-and-pencil exercises completed in one day and addressed art content knowledge and art pedagogical content knowledge. Although the structural scoring approach was used to score both the portfolio entries and assessment center exercises, only the portfolio entries were scored with the software we developed. The rest of this paper will focus, then, on the portfolio entries.

We chose a large with extensive conference facilities, including a sophisticated power system with separate generators and wiring capable of carrying the voltage loads required to supply the seven scoring rooms with power for up to twenty-four computers and videotape players in some instances.

Although the computers could have been networked to the central computer to have information on the assessors' performance virtually on a minute-to-minute basis, that feature would have increased the cost of the scoring sessions. In addition, any problem in the network (for example, a system breakdown) could have jeopardized the flow of the entire scoring sessions and we could not afford taking the risk of loosing information. By not having the computers networked we insured that a problem in any of the computers would not affect the work of the assessors that were working with the other computers.

The price for this decision was that the information on the assessors' performance had to be copied to diskettes from each computer, then fed manually into the central computer. Although labor-intensive, this approach allowed us to update the information on each assessor's performance twice a day.

Participants And Training

127 art teachers acted as assessors of the portfolio responses--an average of 16 assessors for each of the eight portfolio entries. Each of this 127 assessors was provided with a computer. Since 116 teacher-candidates submitted their responses, and the portfolio consisted of 8 entries, a total of 928 portfolio entry responses were scored.[2]

During the first and a half days of the scoring sessions, the assessors were trained on the scoring process, the NBPTS's standards--whose content ultimately determined the content of the assessment, and the use of the scoring rubrics. As a

---

[2] This number is approximate, since some of the responses had been scored previously to be used as "benchmark" responses or "training samples."

part of this training, the assessors scored "training samples" that had been selected and scored previously.

Before "live scoring"--the scoring of real responses, the assessors were also trained on the use PerformAce™. This training took no longer than 30 minutes. Assessors were instructed to score all cases at both the attribute and holistic level.

For each scoring portfolio entry, there was a scoring leader who trained the assessors on the process of scoring, oversaw their performance, provided them with feedback on the quality of their scoring, and decided if they should be re-calibrated or dismissed. There was also an assistant to the scoring leader and a data manager whose functions included handling the candidates' responses to the assessors, receiving from them the responses that had been scored together with the completed scoring forms, and keeping a scoring log to record, among other things, the candidates' responses scored by each assessor, and the holistic scores they have given to those responses.

Three staff members were trained as scoring advisors to interpret the monitoring forms (Figure 4) printed by the central computer and to provide the scoring leaders with updated information on the scoring quality of their assessors based on that information.

Scoring Sessions

The portfolio responses were scored in three and a half days. Each of the responses was scored independently by two assessors. When the difference between the holistic scores assigned by these assessors was equal or greater than 1 point, the response was scored by a third assessor.

To keep a paper backup of the information stored electronically, the scoring procedure required assessors to record the attribute and holistic scores in both the paper scoring forms and the computer.

Each examinee and assessor was assigned an ID number. The assessor had to enter the ID number of the examinee before scoring his or her response. Since each assessor used consistently the same computer, the assessor's ID number was entered automatically. The scores for a given response were thus stored together with the ID number of the candidate and the ID number of the assessor.

Beginning in the middle of the first scoring day, we began printing with the central computer a monitoring form for each assessor. Based on the updated information provided by these monitoring report forms, the scoring advisors met with the scoring leaders twice a day on an individual basis to provide them with specific information on their assessors quality of scoring. This information was given in plain English. For example: "Based on the six cases scored by Assessor 4, and comparing her scores with scores given by other assessors to the same case, it looks like Assessor 4 tends to assign high scores." The scoring leaders could use this information in combination with their own observations to provide feedback to the assessors and to decide when an assessor needed re-calibration.

Score Use

Consistent with the NBPTS' scoring policy, the examinees' responses were scored independently by two assessors on a 12-point scale: 1-, 1, 1+, 2-, 2, 2+, 3-, 3, 3+, 4-, 4, 4+. (For computation and analysis purposes, this scale was translated into: .75, 1, 1.25, 1.75, 2, 2.25, 2.75, 3, 3.25, 3.75, 4, 4.25.) The responses were scored holistically with this scale. The holistic scores were used to make operational decisions (e.g., which responses needed to be scored by a third reader) and certification decisions.

The responses were also scored attribute by attribute with the same scale. The attribute scores were used only for monitoring purposes; they were analyzed to print the monitoring forms.

Instrumentation

At the end of the scoring sessions the assessors and the scoring leaders completed, respectively, an 8-item and an 11-item questionnaire on their satisfaction with PerformAce™ and the scoring approach.

Results

Effectiveness

According to the data provided by the NBPTS's Technical Analysis Group--responsible of examining the psychometric soundness of the assessments developed for the NBPTS, the best possible estimate of the reliability of the assessment was .81, among the highest that have been obtained for any National Board assessment (Jaeger, 1996). Using computers did not affect the assessments' reliability.

Regarding the structural scoring approach and the usability of PerformAce™, we found that: (1) 94% of the assessors agreed that the structural approach helped them to make informed decisions on the quality of teacher performance; (2) 98% agreed that learning to score with the computer was easy, notwithstanding that 47% of the assessors identified themselves as computer-illiterate before coming to the scoring sessions; (3) 73% of the scoring leaders agreed that the information obtained from the scoring advisors allowed them to take specific and effective actions to insure the quality of the scoring, and (4) 100% of the scoring leaders considered that using the information provided by the monitoring form was a "better" or "much better" approach to evaluating and monitoring scoring quality than other traditional approaches such as "reading behind" the cases scored by assessors.

During the first day of "live scoring" we realized that three variables included in the monitoring form were useless or difficult to interpret (see Figure 4). First, the scoring time, which the computer started to count right after the assessor entered the candidate's ID number, was useless because some assessors entered the candidates' ID numbers when they started to review the responses, whereas others

entered those ID numbers after they had reviewed the responses and scored them on the paper scoring form. Second, the layout of the variable, "Adjudicated" was confusing; for example, the zero under the heading, "No" for candidate 148 is a double negation ("it is not true that the response was not adjudicated"). Third, the variable, "Adjudication won?" was not always printed in the monitoring form due to a limitation in the program used by the central computer. We decided to ask the scoring advisors to ignore those variables which, indeed, were not essential to providing information to scoring leaders on their assessors.

Efficiency

Considering the magnitude of the project, the cost of renting the computers used by the assessors in the scoring sessions was low, it did not exceed 1% of the total cost of development of the assessment.

Because the scores were recorded in both paper scoring forms and the computers--which implied more work for the assessors, we cannot draw a final conclusion on whether scoring with the aid of computers reduces scoring time. All we can conclude is that, despite the extra work involved in scoring in both the paper scoring forms and the computers, the scoring sessions ended on time and, in the case of one portfolio entry, several hours earlier than scheduled.

A series of incidents that we faced during the scoring sessions made us identify some issues that are important to implementing computer-assisted scoring. Although the staff members took care of those incidents and none threatened seriously the completion of the scoring, we think that the lessons we learned should be documented.

Table 1 lists those incidents. Incident 1 occurred several times; it resulted from the fact that the facilities used were not designed to connect many computers and other electrical appliances (for example, videotape players) properly; therefore, three or four computers had to be plugged to the same power outlet. Incident 2 is

related to the dependability of the equipment, which we had anticipated--we rented some spare computers. Although score information was always recovered, scoring time was wasted when incidents 1 and 2 occurred, because the assessors who were working with the computers affected could not use them until some staff member fixed the problem.

Incident 3 occurred three times. To make possible that assessors switched places, it was necessary to modify the program of their computers, so their correct ID numbers were attached to the information on the candidates they would score.

Incident 4 occurred because we underestimated the number of computers that would break down. When we ran out of spare computers, we had assessors share computers in pairs: one performed the scoring with the computer while the other was reviewing the response. Three computers were shared this way. We had to modify the program in those computers so that, each time an assessor used it, he or she could enter his or her ID number.

Scoring time did not seem affected when two assessors shared a computer. However, having those assessors enter their ID numbers each time they would score a candidate's response made Incident 5 to occur. A few times, those assessors mistyped their ID numbers.

Although the information on the candidates' scores was never lost due to incidents 1, 2, or 3, tackling those problems was time consuming, we had to distract some of the staff time that could have been better used otherwise. Incidents 5 and 6 never went undetected, but we had to invest a considerable amount of time and work to deal with them: we had to reconcile the computer records with the information in the paper scoring forms and the scoring logs--where the scoring managers kept a record of the responses scored by each assessor. As a result of Incidents 5 or 6, the monitoring form did not print any data for some of the assessors who shared computers.

Conclusions

In terms of effectiveness, the results suggest, first, that a computer-assisted scoring approach does not have an adverse impact on interrater reliability; second, that computers can be used as effective tools to manage scoring sessions by providing valuable and precise information on the quality of the scoring; and third, that appropriate, user-friendly software can be designed so that even computer-illiterate assessors have no problem using it.

In terms of efficiency, the results suggest, first, that implementing a computer-assisted scoring system to score portfolios is not excessively costly; second, that is reasonable to expect that the scoring time is reduced or at least not increased, when scoring is computer-assisted; and third, that the incidents we encountered during the scoring sessions can be readily avoided with some improvements in the planning and the organization of the scoring sessions.

So, in response to the question in the title of this paper--"Can technology enhance the process of scoring portfolios?", the results seem to suggest that the answer is, yes. However, the software and the design of the monitoring form needs to be improved and, the logistics of the scoring sessions needs to be perfected, before assessors can use computers to score portfolios without keeping paper scoring forms as backups. Much better results in portfolio computer-assisted scoring will be obtained in the future if the facilities used for the scoring are properly equipped, so the scoring sessions are not vulnerable to unanticipated incidents.

References

Baratz-Snowden, j. (1991). Performance assessment for identifying excellent teachers. The National Board for Professional Teaching Standards charts its research and development course. Journal of Personnel Evaluation in Education, 5, 133-145.

Cedeño, M. L., & Ruiz-Primo, M. A. (1982). A strategy for assessing methodological and conceptual skills. Unpublished thesis. Mexico: National University of Mexico.

Fitzpatrick, R., & Morrison, E.J. (1971). Performance and product evaluation. In R.L. Thorndike (Ed.), Educational measurement, (pp. 237-270). Washington, D.C.: American Council on Education.

Haertel, E. H. (1990). Performance tests, simulations, and other methods. In J. Millman & L. Darling-Hammond (Eds.). The new handbook of teacher evaluation. Assessing elementary and secondary teachers (pp. 278-294). Newbury Park, CA: Sage Publications.

Jaeger, R. (1996). Conclusions on the Technical Measurement Quality of the 1995-96 Field Test Version of the National Board for Professional Teaching Standards' Early Adolescence Through Young Adulthood/Art Assessment. Technical Analysis Group, Center for Educational Research and Evaluation, University Of North Carolina, Greensboro. September.

Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. Educational Measurement: Issues and Practice, 13(3), 5-16.

Reckase, M. D. (1995). Portfolio assessment: a theroretical estimate of score reliability. Educational Measurement: Issues and Practice, 14(1), 12-14.

Schneider, S., & Austin, S. (1994). Amended proposal for assessment development laboratory: Science. response to RFP #9 questions. San Francisco: Far West Laboratory.

Shulman, L. (1989). The paradox of teaching assessment. In Educational Testing Service (Ed.) New directions for teacher assessment: Proceedings of the 1988 ETS Invitational Conference (pp. 13-27). Princeton, NJ: Educational Testing Service.

Wigdor, A. K., & Green, B. F., Jr. (1991). Performance assessment for the workplace (Vol. 1). Washington, DC: National Academy Press.

Wolf, K. P. (1991). The schoolteacher's portfolio: Issues in design, implementation, and evaluation. Phi Delta Kappan, 73(2), 129-136.

Wolf, K. P. (1994). Teaching portfolios: Capturing the complexities of teaching. In L. Ingvarson & R. Chadbourne (Eds.) Valuing teachers' work. New directions in teacher appraisal (pp. 112-136). ACER.

Table 1. Incidents That Occurred During The Scoring Sessions.

Case 1: Someone kicks accidentally a power outlet and turns off the three or four computers that are plugged to it.
Case 2: A computer breaks down.
Case 3: An assessor is not comfortable in her work station (for example, an air jet is bothering her) and switch places and computers with another assessor.
Case 4: The are no spare computers available, a computer breaks down and cannot be replaced, and the assessor has to share computer with another assessor.
Case 5: When two assessors have to share a computer (see Case 4) an assessor enters a wrong assessor ID number.
Case 6: An assessor enters the wrong candidate's ID number.

Attributes:	1 (lowest quality)	2	3	4 (highest quality)
Scope of Goal	The goal selected is narrow in scope, limited to the teaching of skills related to the process of making works of art.	The goal selected is limited to the teaching of concepts and skills directly related to the process of making, evaluating, and/or interpreting works of art.	The goal selected extends beyond the teaching of concepts and skills directly related to the process of making, evaluating, and/or interpreting works of art.	The goal selected is far reaching in scope, extending beyond the teaching of specific skills and concepts related to the process of making, interpreting, and/or evaluating works of art. It cuts across instructional activities and units and takes a relatively substantial period of time for students to fully understand and internalize.

Figure 1. Description of performance on an attribute at four levels of quality.

Attributes:	1	2	3	4
1. Scope of Goal	The goal selected is narrow in scope, limited to the teaching of skills related to the process of making works of art.	The goal selected is limited to the teaching of concepts and skills directly related to the process of making, evaluating, and/or interpreting works of art.	The goal selected extends beyond the teaching of concepts and skills directly related to the process of making, evaluating, and/or interpreting works of art.	The goal selected is far reaching in scope, extending beyond the teaching of specific skills and concepts related to the process of making, interpreting, and/or evaluating works of art. It cuts across instructional activities and units and takes a relatively substantial period of time for students to fully understand and internalize.
2. Rationale for Goal	Offers little or no rationale for why the goal is important or appropriate for students.	Offers a partial rationale for why the goal is important, but does not fully consider why it is appropriate for students.	Offers a solid rationale for why the goal is important and appropriate for students.	Articulates a clear, consistent and well-thought out rationale for why the goal is important and appropriate for students.
3. Instructional Opportunities Linking to Goal	Instructional opportunities are unrelated to the goal selected.	Instructional opportunities are somewhat linked to the goal selected but they are not clearly defined.	Instructional opportunities are linked to the goal selected and clearly defined.	Instructional opportunities are consistently and insightfully linked to the goal selected and clearly defined.
4. Meeting Student Needs	Teacher offers little or no explanation of how instructional opportunities work together or are related to the goal.	Teacher partially explains how instructional opportunities work together or are related to the goal.	Teacher explains how instructional opportunities work together to meet student needs and to help students make progress toward the goal. Teacher attempts to make connections between instructional opportunities and the selected goal.	Teacher clearly, consistently and insightfully explains in-depth how instructional opportunities work together to meet student needs and to help students make progress toward the goal. Teacher makes students aware of the connections between instructional opportunities and the selected goal.
5. Developmental Understanding of Students	Instructional opportunities reflect little or no knowledge or consideration of adolescents' cultural, artistic, social and intellectual development.	Instructional opportunities reflect basic knowledge and consideration of adolescents' cultural, artistic, social and intellectual development.	Instructional opportunities reflect knowledge and consideration of adolescents' cultural, artistic, social and intellectual development.	Instructional opportunities reflect in-depth knowledge and full consideration of adolescents' cultural, artistic, social and intellectual development.

Figure 2 . Matrix arrangement of the description of performance on several attributes at four levels of quality.

BEST COPY AVAILABLE

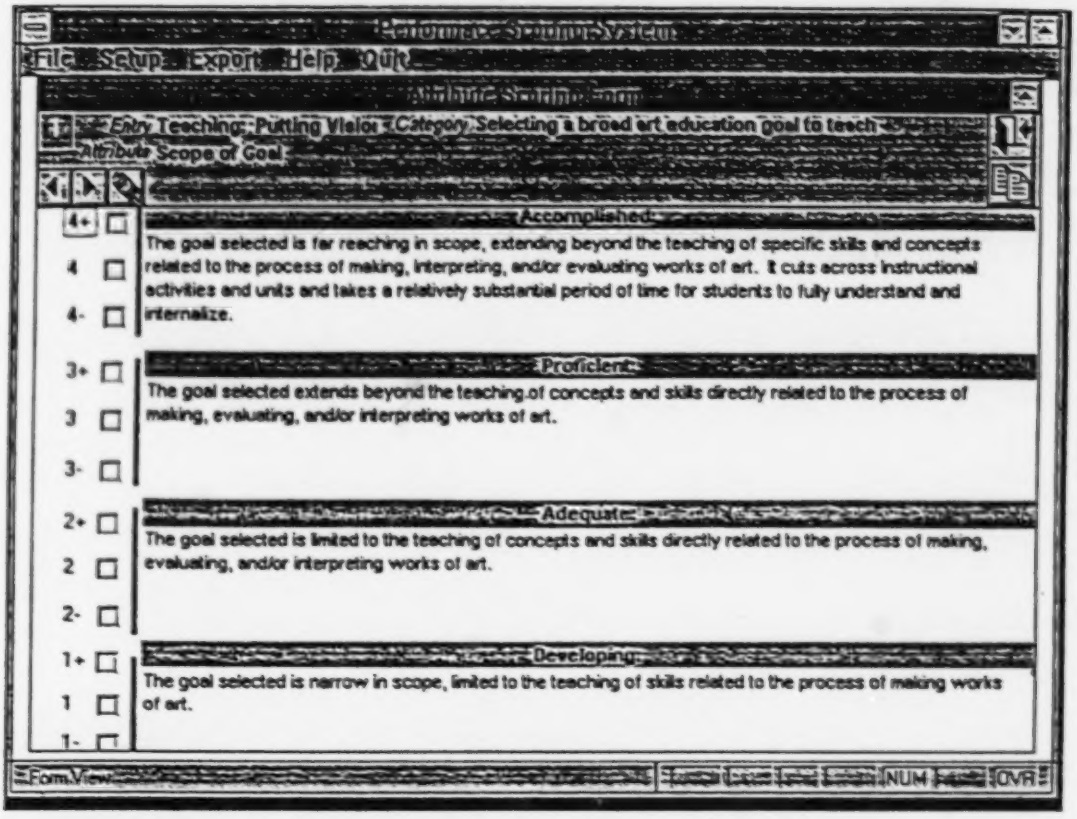
File Setup Export Help Quit
Entry Teaching: Putting Vision Category Selecting a broad art education goal to teach
Attribute Scope of Goal
4+
4
4-
Accomplished
The goal selected is far reaching in scope, extending beyond the teaching of specific skills and concepts related to the process of making, interpreting, and/or evaluating works of art. It cuts across instructional activities and units and takes a relatively substantial period of time for students to fully understand and internalize.
3+
3
3-
Proficient
The goal selected extends beyond the teaching of concepts and skills directly related to the process of making, evaluating, and/or interpreting works of art.
2+
2
2-
Adequate
The goal selected is limited to the teaching of concepts and skills directly related to the process of making, evaluating, and/or interpreting works of art.
1+
1
1-
Developing
The goal selected is narrow in scope, limited to the teaching of skills related to the process of making works of art.
FormView
NUM
OVR

Figure 3. Example of a "screen" displayed on the scorer's computer monitor for one of the attributes.

Monitoring Form

Friday, July 12, 1996 *3:08 PM* Page 7 of 10

Assessor	Entr	Candidate	SeqNo	Time AttributeTotal	Stru Score	Stdv	Corr. with B	HolisticScore Given		Adjudicated No	Yes	Won	Scores 1	2	3
77	1	279	2	175.0	2.21	0.636	0.684	2	2.00	1	0	0	2.75	2	
	1	313	2	130.0	2.57	0.813	0.621	3	3.00	1	0	0	2.75	3	
	1	345	2	121.0	1.93	0.673	0.730	2-	1.75	1	0	0	2	1.75	
	1	377	1	109.0	1.82	0.572	0.702	2-	1.75	1	0	0	1.75	2	
	1	433	2	62.0	1.18	0.278	-0.103	1	1.00	1	0	0	1	1	
	1	434	1	93.0	1.93	0.826	0.889	2-	1.75	1	0	0	1.75	1.25	
	1	474	1	129.0	4.07	0.189	0.702	4	4.00	1	0	0	4	4	
	1	487	1	216.0	2.57	0.572	0.364	3-	2.75	1	0	0	2.75	3	

Figure 4. Assessor monitoring form.

BEST COPY AVAILABLE